ANALYSIS OF BUYING BEHAVIOUR OF CUSTOMERS IN SUPERMARKETS USING DATA MINING TRFM MODEL

AARATI JOSHI, VIPUL VEKARIYA & DAXA VEKARIYA

Department of Computer Engineering, Noble Group of Institutions, Gujarat, India

ABSTRACT

In today's competitive world good marketing strategy is needed to attract the customers. This proposed system maintains the customer relationship using data mining with TRFM model. Clustering, Classification and Association rule are also used with TRFM model that is useful for market intelligence. Clustering is used to search out customer segments with comparable TRFM values. Classification is used to find out customer's future buying pattern. Association rule mining is used for product recommendation.

KEYWORDS: Association Rule Mining, C4.5 Algorithm, Classification, Clustering, Data Mining, Data Preprocessing, FP-Growth Algorithm, K-Means Algorithm, TRFM Model

INTRODUCTION

Supermarkets have huge amount of data of customers in their database. This data is very complicated to search out customer's requirement, their buying pattern. Analysis of Customer's data is useful for making future marketing strategy for Supermarkets.

The method of customer value analysis give the knowledge of future purchasing behaviour of customers from their past buying records. So, TRFM analysis is used to get the customers value in future. In TRFM model four parameters time, recency, frequency, monetary are used. These parameters are used in clustering stage for find out similar customers. After that classification phase occurs, in this phase classification is generated using demographic variables (sex, age, education etc.) of customers. Then finally association rule mining phase comes, this phase is used to find out product recommendation to customers.

RELATED WORK

In [1] authors used simulation model for analysis of buying behaviour of customers in supermarket. Simulation model gives the knowledge about what customer ultimatively do inside the store, namely moving, picking products, buying or not buying.

In [2] author used RFM model for classification of VIP customers. In RFM model, customer rank could be given by three parameters – CP, IP and VIP. This study is very helpful for making useful marketing programs for different customers.

In [3] authors used RFM analysis to get the customers value in future. RFM analysis helps to improve relationship with customers.

In [4] authors used association rule to mine the trusted customers in a supermarkets industry. Association rule

gives the knowledge about which products are together purchased by customer.

In [5] authors used marketing strategy to know about consumer behaviour. Customer behaviour is difficult to predict because of many variables involved and their tendency to correlate with. So, marketing strategy to customer behaviour is made to beat cut throat competition in globally.
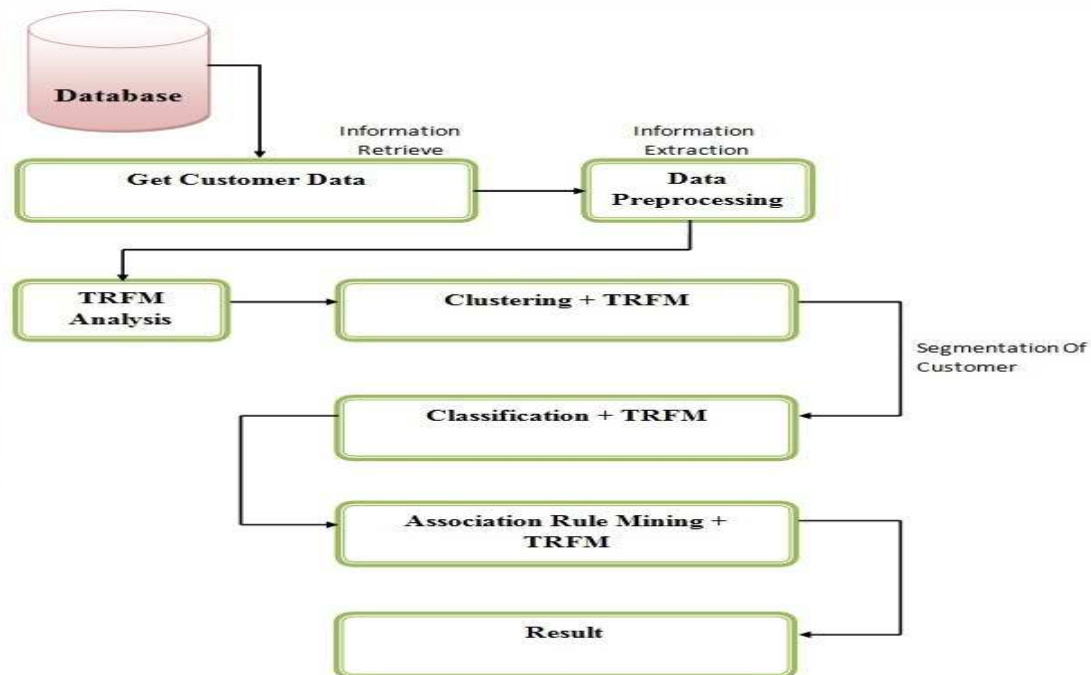
PROPOSED METHODOLOGY

Figure 1: Diagram of Proposed Model

Description of the Proposed Model

Aim of the proposed model is to find out best customer with more accuracy. The proposed model is consists of following steps.

Step 1: Get Customer Data:

First step is to retrieve Customers data from database of supermarkets.

Step 2: Data Preprocessing:

This step is needed to remove missing values, deleting unnecessary attributes, handling outliers and inaccurate values, transforming data into proper format, discretizing the original values into small number of value ranges, replacing low level concepts by high level concepts. After data preprocessing information of customer's data would be extracted.

Step 3: TRFM Analysis:

In TRFM analysis four parameters- Time, Recency, Frequency, Monetary is considered.

Time (T): Time (T) gives the information about how much time taken by consumer to buy products.

Recency (R): Recency (R) gives the information about how recent was the last purchase of products by consumer.

Frequency (F): Frequency (F) gives the information about how often does the consumer purchase the products.

Monetary (M): Monetary (M) gives the information about how much money spent by consumer to purchase products.

TRFM Analysis would be done in Three Steps Introduced in the Following

Step 3.1: In this step rank of customers would be given on the basis of T-R-F-M attributes. Partition of four T-R-F-M attributes into 5 equal parts and each part is equal to 20% of all. By taking following sample, this step would be easily understood:

Table 1:Time Ranking would be Given as Follows

Time Ranking	Time (Time taken by customer for product purchases)
5	Less than 289 minutes
4	289 - 576 minutes
3	577 - 864 minutes
2	865 - 1152 minutes
1	1153 – 1440 minutes

Table 2: Recency Ranking would be Given as Follows

Recency Ranking	Recency (Days since last purchases)
5	Less than 74 days
4	74 – 146 days
3	147 - 219 days
2	220 - 292 days
1	293 - 365 days

Table 3: Frequency Ranking would be given as Follows

Frequency Ranking	Frequency (Number of shopping)
5	More than 7 times
4	6 – 7 times
3	4 – 5 times
2	2 - 3 times
1	Once

Table 4: Monetary Ranking Would be given as Follows

Monetary Ranking	Monetary (Amount spent)
5	More than $7999
4	$6000 - $7999
3	$4000 - $5999
2	$2000 - $3999
1	Less than $2000

Step 3.2: In this step TRFM score would be generate. TRFM score of above sample would be as follows:

Table 5

TRFM cell score	Time	Recency	Frequency	Monetary
5555	250 minutes	50 days	11 times	$9000
3441	600 minutes	130 days	6 times	$800
4354	400 minutes	165 days	10 times	$7000
2232	1000 minutes	230 days	5 times	$3000
1111	1200 minutes	360 days	Once	$500

Here, in this sample only 5 TRFM cell score would be taken. There are total 625(5 x 5 x 5 x 5 = 625) combination of each attribute in T-R-F-M attributes would be generate.

Step 3.3: In this step, TRFM cell score for all the customers would be sort either ascending or descending order. Customers with TRFM score as 5555 are best customers with those with 1111 are the least desirable customers. After this step best customers would be identified. Best customers are those who increase the profit of the supermarkets.

Step 4: Clustering + TRFM:

There are many clustering techniques but here K-means algorithm is used for clustering of TRFM values. Basic steps of K-means algorithm are as follows [6]:

Let $X = \{x_1, x_2, x_3, \ldots\ldots, x_n\}$ be the set of data points and $V = \{v_1, v_2, \ldots\ldots, v_c\}$ be the set of centers.

Step 1: First decide the number of clusters.

Step 2: Then, randomly select cluster centers 'c'.

Step 3: Now calculate the gap between each data point and cluster centers.

Step 4: Assign the data point to the cluster center whose distance from the cluster center is nearest of all the cluster centers.

Step 5: Recalculate the new cluster center using:

$$v_i = \left(\frac{1}{c_i}\right) \sum_{i=1}^{c_i} x_i$$ Where, 'c_i' denotes the number of data points in i^{th} cluster.

Step 6: Again calculate the distance between each data point and new achieved cluster centers.

Step 7: If there is none of data point was reassigned then stop, otherwise repeat from step 4.

Here, clustering is used to partition large data sets of customers into groups according to their similar TRFM values. This step is useful for making variant marketing strategy for variant customer segments.

Step 5: Classification + TRFM:

Classification rule are generated on the bases of demographic variables (sex, age, education etc.) and result of step 4. There are many classification techniques but here C4.5 decision tree algorithm is used for classification. Basic steps of C4.5 algorithm are as follows [8]:

Step 1: Check for base cases.

Step 2: For each attribute a calculate: Normalized information gain from dividing on attribute a.

Step 3: Select the best a, attribute that has highest information gain.

Step 4: Create a decision node that divides on best of a, as root node.

Step 5: Recurs on the sub lists achieved by splitting on best of a and add those nodes as children node.

After this step customer would be classify according to their profile information related to sex, age, marital status, education. So, it would be helpful to market manager for better understand of the customer's data at large.

Step 6: Association Rule Mining + TRFM:

The motivation behind this step is to classify the association between consumer segments, consumer profiles and product items purchased together. Here FP-Growth algorithm is used for association rule mining. Major steps of FP-Growth algorithm are as follows [10]:

Step 1: It firstly compresses the database showing frequent item set in to FP-tree. FP-tree is built using 2 passes over the dataset.

Step 2: : It splits the FP-tree in to a set of conditional database and mines each database separately, so extract frequent item sets from FP-tree directly.

In FP-growth algorithm, FP-tree would be constructing after step 1 and after step 2 frequent items set would be generated.

RESULTS

This step gives the knowledge of the proposed system.

CONCLUSIONS

RFM analysis differentiate important consumers from huge database by three parameters as interval of recency, frequency and monetary. But how much time taken by customer to purchase the products in supermarket is not considered in RFM analysis. So, TRFM analysis will give the knowledge of customers time period of purchase products. Customer with an overall high TRFM score represents the best customer. With use of TRFM, Customers who take less time to purchase products are more desirable customers than Customers who take more time to purchase products. In the competing world of today, TRFM analysis helps supermarkets to better attain their goals of profit and customer relationship. In this way, TRFM score will give the best customer by more accuracy than RFM analysis.

REFERENCES

1. Clemens Schwenke, Volodymyr Vasyutynskyy and Klaus Kabitzsch, 'Simulation and Analysis of Buying Behavior in Supermarkets', *IEEE Conference on Emerging Technology and Factory Automation (ETFA)*, pp. 1-4, 2010.

2. Wei Jainping, 'Research on VIP Customer Classification Rule Base on RFM Model', *International Conference on Management Science and Industrial Engineering (MSIE)*, pp. 336-338, 2011.

3. Divya D. Nimbalkar and Asst Prof. Paul, 'Data mining using RFM Analysis', International *Journal of Scientific & Engineering Research*, Vol. 4, Issue 12, pp. 940-943, 2013.

4. Abhijit Raorane and R.V.Kulkarni, 'DATA MINING TECHNIQUES: A SOURCE FOR CONSUMER BEHAVIOR ANALYSIS', *International Journal of Database Management Systems (IJDMS)*, Vol. 3, No. 3, pp. 45-56, 2011.

5. Sunanda Sharma and Dr. Kashmiri Lal, 'CHANGING CONSUMER BEHAVIOUR- A CHALLENGE FOR SUSTAINABLE BUSINESS GROWTH', *International Journal of Marketing, Financial Services & Management Research*, Vol. 1, Issue 8, pp. 149-158, 2012.

6. Archana Singh, Avantika Yadav and Ajay Rana, 'K-means with Three different Distance Metrics', *International Journal of Computer Applications*, Vol. 67, No. 10, pp. 13-17, 2013.

7. Tapas Kanungo, David M. Mount, Nathan S. Nethanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu, 'An Efficient k-Means Clustering Algorithm: Analysis and Implementation', *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, Vol. 24, No. 7, pp. 881-892, 2002.

8. Gaurav L. Agrawal and Prof. Hitesh Gupta, 'Optimization of C4.5 Decision Tree Algorithm for Data Mining Application', *International Journal of Emerging Technology and Advanced Engineering*, Vol. 3, Issue 3, pp. 341-345, 2013.

9. Salvatore Ruggieri, 'Efficient C4.5', *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, Vol. 14, No. 2, pp. 438-444, 2002.

10. Kuldeep Malik, Neeraj Raheja and Puneet Garg, 'ENHANCED FP-GROWTH ALGORITHM', IJCEM *International Journal of Computational Engineering & Management*, Vol. 12, pp. 54-56, 2011.

11. Pratiksha Shendge and Tina Gupta, 'Comparitive Study of Apriori & FP Growth Algorithms', *PARIPEX – INDIAN JOURNAL OF RESEARCH*, Vol. 2, Issue 3, pp. 20-22, 2013.